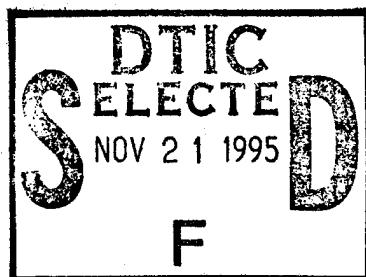


IDA PAPER P-3076

## A MODEL OF FALSE ALARMS IN TARGET ACQUISITION BY HUMAN OBSERVERS



James D. Silk

September 1995

19951117 035

*Prepared for*  
Advanced Research Projects Agency

Approved for public release; distribution unlimited.

DTIC QUALITY INSPECTED 5

INSTITUTE FOR DEFENSE ANALYSES  
1801 N. Beauregard Street, Alexandria, Virginia 22311-1772



## **DEFINITIONS**

**IDA publishes the following documents to report the results of its work.**

### **Reports**

Reports are the most authoritative and most carefully considered products IDA publishes. They normally embody results of major projects which (a) have a direct bearing on decisions affecting major programs, (b) address issues of significant concern to the Executive Branch, the Congress and/or the public, or (c) address issues that have significant economic implications. IDA Reports are reviewed by outside panels of experts to ensure their high quality and relevance to the problems studied, and they are released by the President of IDA.

### **Group Reports**

Group Reports record the findings and results of IDA established working groups and panels composed of senior individuals addressing major issues which otherwise would be the subject of an IDA Report. IDA Group Reports are reviewed by the senior individuals responsible for the project and others as selected by IDA to ensure their high quality and relevance to the problems studied, and are released by the President of IDA.

### **Papers**

Papers, also authoritative and carefully considered products of IDA, address studies that are narrower in scope than those covered in Reports. IDA Papers are reviewed to ensure that they meet the high standards expected of refereed papers in professional journals or formal Agency reports.

### **Documents**

IDA Documents are used for the convenience of the sponsors or the analysts (a) to record substantive work done in quick reaction studies, (b) to record the proceedings of conferences and meetings, (c) to make available preliminary and tentative results of analyses, (d) to record data developed in the course of an investigation, or (e) to forward information that is essentially unanalyzed and unevaluated. The review of IDA Documents is suited to their content and intended use.

The work reported in this document was conducted under contract DASW01 94 C 0054 for the Department of Defense. The publication of this IDA document does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that Agency.

IDA PAPER P-3076

## A MODEL OF FALSE ALARMS IN TARGET ACQUISITION BY HUMAN OBSERVERS

James D. Silk

September 1995

Approved for public release; distribution unlimited.



INSTITUTE FOR DEFENSE ANALYSES

Contract DASW01 94 C 0054  
ARPA Assignment A-162

## PREFACE

This paper has been prepared for Mr. Thomas Hafer, Deputy Director Advanced Systems Technology Office, ARPA, in partial fulfilment of an IDA task order on Analysis and Model Development. Additional cognizance and direction have been provided by Mr. John Brand and Mr. Eugene Patrick, U.S. Army Research Laboratory (ARL), S<sup>3</sup>I Special Projects Office; and Mr. John D'Agostino, U.S. Army Night Vision and Electro-optics Systems Directorate (NVESD), Visionics Division.

These analyses would not have been possible without the high quality target acquisition performance data obtained by the Visionics Division of NVESD in their Phase I and Phase IV target acquisition tests.

Accesion For	
NTIS	CRA&N
DTIC	TAB
Unannounced	
Justification	
By _____	
Distribution / _____	
Availability Codes	
Dist	Avail and / or Special
A-1	

## CONTENTS

EXECUTIVE SUMMARY .....	S-1
I. INTRODUCTION .....	1
A. Background .....	1
B. Purpose and Activities .....	1
C. Findings .....	2
II. THE OBSERVER TESTS .....	5
III. OBSERVER EXPECTATIONS .....	7
IV. TRENDS IN THE TEST RESULTS .....	11
V. MODEL DEVELOPMENT .....	15
VI. MODEL FITS AND POPULATION STATISTICS .....	21
VII. FALSE ALARM PREDICTION FROM IMAGE STATISTICS .....	29
VIII. SUMMARY .....	33
Appendix A—Observer Test Data .....	A-1
Appendix B—False Alarm Location Analysis .....	B-1
Appendix C—Transformation of Averaged Variables .....	C-1

## FIGURES

III-1. Average detection and false alarm rates from a subset of Phase IV data .....	8
III-2. As Figure III-1, but for Phase I.....	8
IV-1. Detection probability vs. false alarm rate by observer for two of the views in the Phase I test b1.....	12
V-1. Detection and false alarm data from Phase I test b1 view 5.....	15
V-2. The $c=1$ curve shows greater sensitivity at $FAR>1$ , but the $c=3$ curve is better for $FAR<1$ .....	17
V-3. Illustration of the definition of the threshold and sensitivity variables .....	18
VI-1. Comparison of the dependence of the slope (or width) parameter, $c$ , on view for the various sensor simulations (panel a, on left) and observer pools (panel b, on right) .....	22
VI-2. Summary statistics of the logodds ( $P_D$ ) distribution for the various views .....	23
VI-3. Frequency distribution of detections for the eight views of test b1 .....	24
VI-4. Standard deviations of the observer populations for the 0% (left) and 100% confidence thresholds .....	25
VI-5. Thresholds for 100% confidence vs. 0% confidence for all views and tests, averaged over observers .....	26
VII-1. Correlation of false alarm rate with statistical variance, for low (left) and high (right) confidence detection .....	26

## TABLES

VI-1. Observer population statistics derived from Phase I Target Acquisition tests .....	27
VII-1. Results of power fits of the form $FAR = \text{constant} \times SV^{\text{power}}$ to the data of Figure VII-1 .....	31

## EXECUTIVE SUMMARY

IDA has conducted a thorough analysis of the false alarm data from the U.S. Army Night Vision and Electro-Optics Systems Directorate perception experiments. IDA's analysis was designed to support model development by addressing specific modeling needs:

1. A predictive model of false alarm performance. For a given clutter environment, how many false alarms are expected on the average?
2. A descriptive model of the observer ensemble. How can variations among observers be quantified?

This work reports on the above effort.

### PREDICTIVE MODEL

The relationship between false alarm rate (FAR) and clutter level is documented for a test in which observers have been conditioned to a target-rich environment. Good correlation exists between a standard measure of clutter and false alarm rate. But the correlations do not persist for observers in a different test, who were conditioned to expect fewer targets. *Observer state turns out to be a much stronger driver of false alarm rate than clutter level.* This observation does not apply to the probability of detection,  $P_D$ .

The IDA false alarm model is valid only for the specific high-target-density scenario of the Phase 1 test. It does not consider variation in the observer state. Until better understanding of this effect is at hand, *IDA recommends against implementing the predictive portion of its false alarm model.* Further testing is needed, but we believe that the required tests are fundamentally unlike those conducted to date.

In tests to date, observer state was pre-conditioned, the learning curve was saturated, and the state of the observer was not varied within a test. All of this is in keeping with commonly accepted practice. Resolving the false alarm issues [and, similarly, identification friend or foe (IFF) and fratricide issues] requires a new generation of tests in which the subjects experience controlled, repeatable preconditioning. Extended, detailed simulations and/or extensively instrumented, realistic exercises may be appropriate.

## DESCRIPTIVE MODEL

We have had better success in characterizing the observer ensemble than the predicted false alarm rate. We have produced a useful approach to explaining the differences that appear among the test subjects.

In any difficult discrimination task, there is a tradeoff between errors of the first and second kinds. For target detection, a missed target is an error of the first kind; a false alarm is an error of the second kind. By being more or less conservative in his declarations—that is, by setting a high or low *threshold*—an observer can trade one kind of error for the other.

On the other hand, some subjects are, through nature or nurture, better observers than others. For example, high visual acuity or familiarity with target features may yield an advantage. This inherent level of discrimination capability is called *sensitivity*. Enhanced sensitivity implies the ability to simultaneously reduce errors of both kinds.

We find from these data that there is much less variation in sensitivity to target/background discrimination than there is in the threshold that the observer sets for himself. We have demonstrated a useful parameterization of these effects and provided numerical support for implementing simulations.

*We believe that the descriptive approach developed herein will be useful to the wargaming community, and we recommend that it be considered as a starting point for a more complete model.*

## I. INTRODUCTION

### A. BACKGROUND

The modeling of target acquisition by humans has focused on the problem of predicting whether targets will be detected. In its simplest form, the target detection problem is to compute the fraction of a standard observer ensemble that will detect a given target. The computation is based upon input about the target state, sensor characteristics, and scenario.

The closely related issue of false alarm prediction has received less attention. There is good reason for this. From a user viewpoint, lethality (for sensors) and vulnerability (for signature control) are more directly affected by true detections than by false ones. Thus, for purposes of evaluating materiel systems, the need for detailed modeling of false alarms is marginal.

While modeling of false alarms is secondary to true detection, it is nevertheless important to understand the effects of false alarms and to account for them in the development of doctrine. For the soldier performing a target acquisition task, the detection of true targets drives his lethality. On the other hand, servicing false targets affects his survivability: His response time, munitions stores, and state of concealment are all compromised. Moreover, the subject is closely linked to collateral damage and fratricide.

### B. PURPOSE AND OBJECTIVES

In this work we extend the scope of target acquisition modeling to the consideration of false detections. The model that we propose establishes a link between false alarm rate and a simple scene complexity statistic. More significantly, it provides a statistical representation of the correlation among the observer ensemble of false alarm rate with the target detection probability.

The model is phenomenological in the sense that it seeks only to describe the results of the tests. It does not build on psychophysical "first principles," but is entirely data driven. The results are tabulated and presented in such a way that they could easily be implemented in the combat simulations by the wargaming community.

The descriptive model is based on the analysis of data obtained in a series of target acquisition tests. The tests were conducted by the Visionics Division of the Night Vision and Electro-Optics Directorate [NVESD]. They were designed (see Section II) to represent (as realistically as possible in a laboratory setting) the target acquisition task faced by a tank gunner in the course of an engagement.

### C. FINDINGS

An important finding from the analysis of the test data is that the dominant determinant of false alarm rate is the *expectation* of the human subject. We show how seemingly subtle aspects of observer expectation can dramatically affect false alarm rates. We present this result early on (and expand on it in Section III) so that the results that follow can be placed in the proper perspective, with adequate caveats. The principle caveat is that the present results should not be extended to target acquisition tasks beyond the one simulated in the tests.

A more general review of the test results reveals two striking features (see Section IV). First, for a given background scene, strong positive correlation is seen among the observers between the average false alarm rate and the average detection probability. Second, when the observers are asked to lower their discrimination threshold from standard military detection to "possible object of interest," the trend in the data is extended seamlessly.

The two cited features of the data strongly suggest a description based on signal detection theory and we construct such a model of the observer responses (Section V). The simplest model based on this paradigm would enable us to characterize the various backgrounds by different values of a single sensitivity parameter. Unfortunately, straightforward application of the theory is not appropriate here. This is so because, in a realistic search test, the number of "true dismisses" is not known. It is therefore impossible to convert the false alarm rate to a false alarm probability. We finesse this issue by introducing the normalization as a second free parameter, but find that the data is not well described by this model. A more general model, obtained by introducing a third parameter, describes the data quite well. Happily, a simple and appropriate approximation collapses one of the constraints on the model, so only two free parameters are really needed.

Re-analysis of the test data in the context of the model allows us to extract the parameters that describe the observer ensemble. We tabulate the results (Section VI) and find that observer threshold varies much more than observer sensitivity. Moreover, the

threshold shift corresponding to the resetting of discrimination threshold from "full detection" to "possible object of interest" is remarkably reproducible across sensors, backgrounds, and observers.

Finally, we demonstrate the correlation between false alarm rate and a scene complexity statistic. We have not accomplished the demonstration of a good correlation between our second free parameter and any scene or target statistic. As an interim measure we therefore estimate its mean and variance parameters so that stochastic simulations can be implemented appropriately.

We include the complete set of displays of the test data (Appendix A). In Appendix B we summarize our previously unpublished analysis of the statistical variance (SV) statistic as a predictor of false alarm locations. Appendix C contains the details of the computation of some average quantities.

## II. THE OBSERVER TESTS

The target acquisition tests that form the basis of this paper were conducted by NVESD as part of the Army's Thermal Target Acquisition Model Improvement Program (TAMIP). More complete descriptions of the tests and the data accrued from them can be found elsewhere.<sup>1</sup> The scope and objectives of the tests go well beyond the application presented herein, and several analyses of other aspects of the tests have already been published.<sup>2</sup> The present description is thus intentionally both incomplete and imprecise. It is intended only to convey sufficient information for the work in hand to be self-contained and comprehensible.

The tests referenced in this paper are designated by NVESD as Phase I tests a1, b1, b2, and b3 (this is the complete set of Phase I tests); and Phase IV test a (this is one of four Phase IV tests). The methodologies of the two Phases are quite similar, but their differences are important. The bulk of our analysis is supported by the Phase I data. The explanation for this preference, along with a brief analysis of Phase IV data, is the subject of Section III. We will first discuss the Phase I tests, then note the relevant differences between these and the Phase IV a test.

In each of the four Phase I tests, each observer was shown the same set of images, or "scenarios," on the terminal screen of a desktop computer. They were instructed to designate, using a mouse, all possible target candidates (or "areas of interest"); then to go back over their selections and assign one of four confidence levels (0, 25, 50, or 100 percent confidence) to each selection. All responses, including incorrect ones, were recorded for offline scoring and analysis.

---

<sup>1</sup> Barbara L. O'Kane, Clarence P. Walters, John D'Agostino, "Report on Perception Experiments in Support of Thermal Performance Models," NVESD Report, February 1993.

<sup>2</sup> Barbara L. O'Kane, Clarence P. Walters, John D'Agostino, Mel Friedman, "Target Signature Metrics Analysis for Performance Modeling," *Proceedings of the IRIS Symposium on Passive Sensors*, Volume 2, p. 161, 1993; John D'Agostino, Russ Moulton, Bob Sendall, Walt Lawson, "MFTD - A Measure of Sensor Performance Under Scene Clutter Limited Conditions," NVESD Report, March 1993; John D'Agostino, "TAMIP Thermal Modeling Program: Final Technical Report for 1993," NVESD Report, May 1994.

The images were derived from photographs of the NVESD terrain board, which is a scale model representative of Central European terrain. Scale models of military vehicles are placed on the terrain board. The terrain and vehicles are painted and photographed so as to resemble infrared signatures as seen through a FLIR sensor. Postprocessing of the digitized imagery introduces specific sensor effects.

Sixty images were shown to each subject. Each image contained from three to seven military targets, or none at all, for a total over all images of 275 targets. The 60 images were divided among eight different backgrounds, corresponding to differing levels of clutter. Each background appeared in seven or eight different images, and each appeared once without targets.

The Phase I b1, b2, and b3 tests were each conducted using the same pool of 22 military observers as subjects. These three tests all used the same basic images, but processed differently to simulate three different sets of sensor characteristics. Roughly speaking, the b2 test corresponds to a low noise, high resolution sensor. The b1 and b3 tests represent lower resolution and higher noise excursions, respectively, from the b2 sensor. The a1 test used the same scenarios and sensor simulation as the b1 test, but the 17 subjects were civilian analysts.

The conduct of Phase IV test a was similar in most respects to the Phase I tests. The principal difference is that the Phase IV imagery is real infrared sensor imagery of real terrain and real military vehicles. There is a greater variety of backgrounds in the Phase IV image set, although there is a subset in which the background is a controlled variable. In this subset, image processing techniques were to alter the signature of the military vehicles while leaving the terrain background unchanged.

As we shall see, another important difference is that there is exactly one target in about 85 percent of the Phase IV test a images.

### III. OBSERVER EXPECTATIONS

Since the image stimuli used in the Phase IV test a are based on real, not modeled, infrared signatures and sensors, it would be preferable to base our modeling effort on this data set rather than the Phase I set. Early in the analysis of the Phase IV data, however, it became clear that the observers learned to expect one target per image. This expectation on the part of the subjects affected performance in a way that we will demonstrate in this section. We believe that the behavior that is reflected in this data is not representative of behavior in the operational environment. We also believe, for reasons that are discussed below, that the cause of this behavior is not present in the Phase I tests. Therefore it is appropriate to use the Phase I data, despite the lower fidelity of its imagery, in the development of our model for false alarms.

To demonstrate the existence and effect of the observer expectations in the Phase IV test a data, we focus on a subset of the images. This subset contains 70 images, which are partitioned into ten subsets. Within each subset of seven images, all the backgrounds are identical; the target alone varies. One of the seven is the baseline, which is the image as obtained in the field. In the other six images, the target pixels have been subjected to a specific treatment so as to make the target look less like a military vehicle. Each subset contains one baseline and one image corresponding to each treatment. The detection probabilities and false alarm rates which we shall now discuss are averages for the baseline targets and the six target treatments, across subsets; that is, each average is taken over all ten backgrounds belonging to a treatment (or baseline), and over all 36 observers.

Figure III-1 shows the effect of the various treatments on the detection probabilities and false alarm rates. It is evident that the treatments were effective to various degrees in increasing the number of times the observers missed the target. It is equally clear that the treatments resulted in an increase in false alarm rate that is strongly correlated with the number of misses. But recall that all of the points in the figure correspond to identically similar backgrounds. Only the target pixels were altered between treatments. The only plausible explanation for the depicted behavior is that subjects responded differently to identical background stimuli based on whether they had detected a target somewhere in the image.

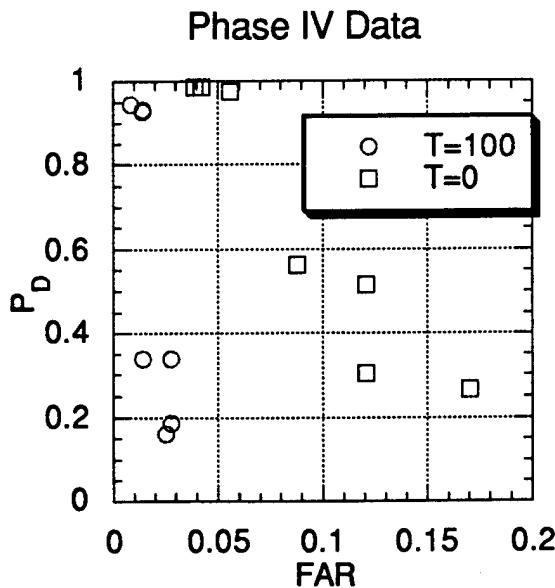


Figure III-1. Average detection and false alarm rates from a subset of Phase IV data. The averaging procedure is defined in the text. The T values refer to observer confidence designations.

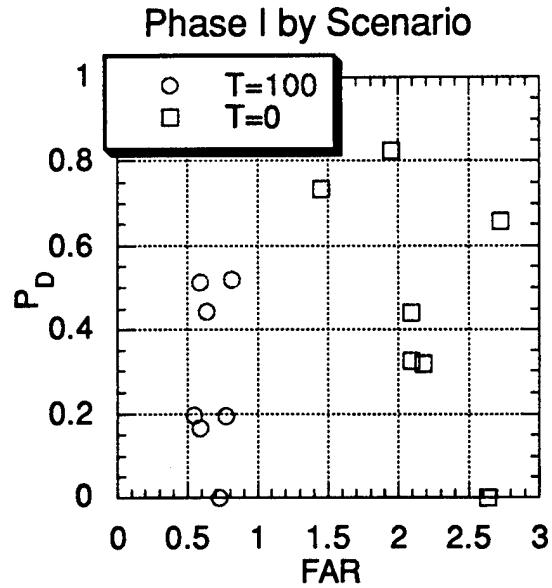


Figure III-2. As Figure III-1, but for Phase I.

Such behavior is at odds with reasonable expectations of what would happen in the operational environment. Certainly observer expectations are equally important there. But it seems more plausible that the expectation of encountering targets would be *enhanced* by the identification of a true target, rather than diminished.<sup>3</sup>

To be sure, the observers expected to find targets in all of the Phase I images as well. We contend, however, that this expectation has a much less confounding effect on the results there than in Phase IV. Since there were so many targets per scenario in those tests, the observers almost always resolved their expectation by detecting a true target. Their expectation did not force them into the false alarm regime.

A look at the Phase I data supports these ideas. Figure III-2 is the Phase I analog to Figure III-1. Again, each point corresponds to identical backgrounds but different targets. While each average in the Phase IV set contained one target and ten backgrounds, the

<sup>3</sup> It is tempting, though, to speculate that even in a realistic scenario a "difficult" target is more likely to be detected if there are no "easy" targets in evidence. If an observer is certain that there are detectable targets nearby, this is obviously so. But even if the observer is uncertain, the "easy" target may create an expectation that other targets will be equally easy. Such are the difficulties in assessing the importance and impact of the state of the observer.

averages in the Phase I set contain several targets (with one exception, see below) and one background. The principal difference is that the various targets *appeared in the same scenario*. We observe that there is much less variation in the number of false alarms among the points in Fig. III-1 than among those in Fig. III-2, even though there is a similar variation in detection probability. Just as important as the lower variation is the lack of correlation with detection probability in the Phase I trend. We therefore accept the explanation that the observer expectation to find targets is, for practical purposes, always resolved in the Phase I test but not in the Phase IV test.

It is important to point out, though, that the sham run in the Phase I test [that is, the scenario without targets (it appears in Fig. III-2 as the scenario with  $P_D = 0$ )] is also in line with the false alarm rates for high  $P_D$ . This seems at odds with our conjecture that it was the high multiplicity of targets in the Phase I tests that would resolve the observer expectations and render the false alarm rate stable. A comparison of the horizontal axes of Figures III-1 and III-2 provides the explanation. The observers in the Phase I test were declaring false alarms in far greater numbers—by a factor of 20—than were the Phase IV observers. In fact, even for the high threshold responses, the Phase I false alarm rates approach or exceed one false alarm per scenario per observer. Thus, the observer's expectation to find a target is, at least to some extent, quenched by the false alarms themselves. But this greater propensity for false alarms is certainly driven predominantly by the observer expectation of multiple targets per scenario, as suggested in the footnote. So, whether directly or indirectly, the higher target density in the Phase I data tends to make the effects of observer expectations less confounding than in the Phase IV test.

None of the above should be taken to imply that the Phase IV data may not yield detection probability data that is useful. While the detection probability measurements are not immune to the effects of observer expectation, we are not concerned that the effect on the analysis of Phase IV detection probabilities will be so troublesome. As we shall see in the following sections, the detection probabilities are far less sensitive to observer expectations than are the false alarm rates.

The foregoing discussion illustrates the limits of validity for the model, which we shall base upon the Phase I data set. Clearly the limits will apply only for very target-rich environments. It is doubtful, though, that the lower limit of validity for target density is as high as several per field of view. In a real search, the observer has the option to return to previously scanned areas; the freedom to revisit did not exist in either phase of the tests. Neither will such freedom apply in rapidly changing environments. The expectations of the

observers in the tests were relatively constant throughout the testing, and in fact had been preconditioned before the scored part of the test in order to saturate the learning curve.

## IV. TRENDS IN THE TEST RESULTS

Our approach to the modeling of false alarms is phenomenological. We seek to identify and parameterize whatever trends exist in the data, not to force them into a preconceived model nor even necessarily to explain them. In this section we illustrate the trends that are apparent upon a rudimentary analysis of the data in order to motivate the model development and more extensive analysis which follow.

All of the qualitative effects discussed in this section hold equally well for all of the four Phase I data sets (a1, b1, b2, and b3). Here we focus on the b1 test data. Final results for all data sets will be displayed and discussed in Section VI and the Appendix.

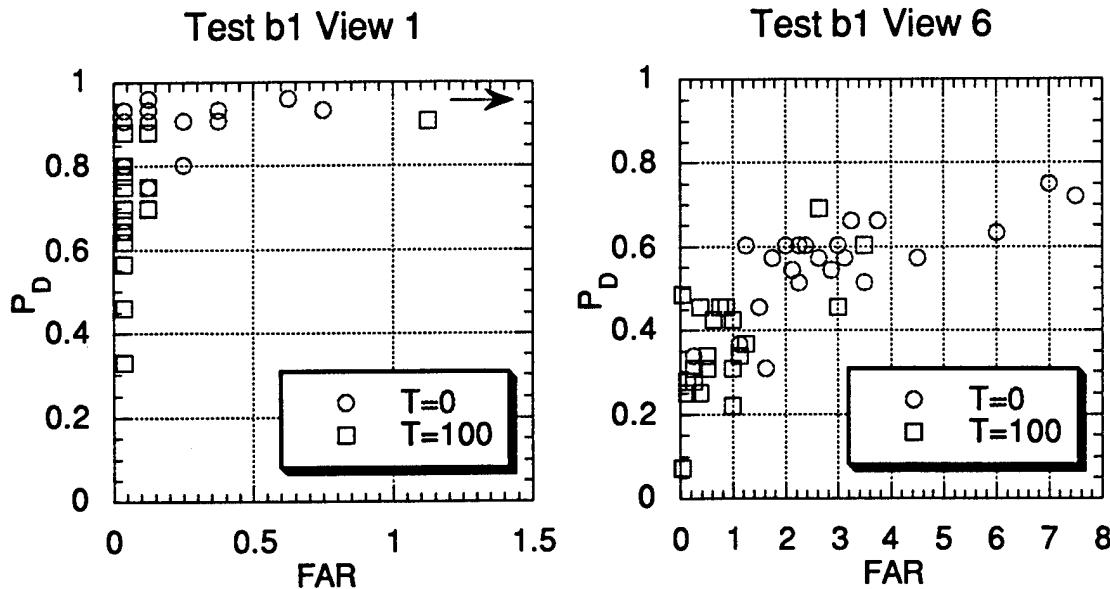
First let us describe the preliminary data processing steps for the sake of clarity and defining terminology. Since we expect false alarm rates to depend strongly on the background scene, we partition the data set into the eight subsets corresponding to the eight background scenes, or "views," used in the test. For each observer, we compute detection probabilities and false alarm rates over all the images, or "scenarios," in that view.

The observers were asked to select a confidence level for each target candidate designation. We shall treat these confidence levels as thresholds, and compute detection probabilities (or false alarm rates) at a given threshold based on the number of correct (or incorrect) declarations with confidence equal to or greater than that threshold. Thus, probability of detection ( $P_D$ ) for  $T = 100$  is always less than or equal to  $P_D$  for  $T = 0$ . Note that NVESD considers the  $T = 100$  threshold to correspond to "full military detection." The  $T = 0$  threshold is said to represent the declaration by the observer of a "possible area of interest."

For computation of probabilities of detection we use one of a family of so-called uninformed estimates.<sup>4</sup> This prescription gives the maximum likelihood estimate of probability for  $n$  occurrences out of  $N$  trials as  $(n+1/2)/(N+1)$ , instead of the more commonly used  $n/N$ . Our motivation is primarily to avoid singularities that would arise if probabilities were allowed to take the values 0 or 1. We choose this prescription over other

---

<sup>4</sup> Harry F. Martz and Ray A. Walker, *Bayesian Reliability Analysis*, John Wiley & Sons, 1982.



**Figure IV-1. Detection probability vs. false alarm rate by observer for two of the views in the Phase I test b1. The arrow in the left panel indicates a  $T = 0$  point beyond the limit, at (4.3, 0.96).**

candidates [for example,  $(n+1)/(N+2)$ ] because the distribution of  $P_D$  that was measured in these tests was "U-shaped" and not uniform.

We define the false alarm rate (FAR) as the number of false alarms per scenario per observer. The false alarm rate for zero occurrences is estimated as if 1/3 of an event had occurred. For example: If observer A has 0, 1, 3, 0, 1, 2, and 0 false alarms in the seven scenarios belonging to view 6, then his FAR is  $7 / 7 = 1$ . If observer B has zero false alarms for all seven scenarios, then his FAR is  $\frac{1}{3} / 7 = 0.048$ .

Figure IV-1 shows representative plots of  $P_D$  vs. FAR for two of the views. In each case, the square symbols correspond to the  $T = 100$  threshold and the circles to  $T = 0$ . We note that the observers fall along relatively well-defined trajectories in the plane. The two trajectories are quite different in the two panels. The left hand panel corresponds to a low clutter background scene, the right to high clutter. It is clear that the low clutter case admits a trajectory that is much closer to the ideal performance (upper left corner, where  $P_D = 1$  and  $FAR = 0$ ) than the high clutter case.

It is particularly interesting that the low threshold data forms an apparently seamless continuation of the high threshold subset. That is, there are some observers whose  $T = 100$  performance closely matches the  $T = 0$  performance of others, both in  $P_D$  and FAR. Thus, zero confidence and 100 percent confidence clearly mean vastly different

things to different observers. While it has long been recognized that some observers are "better" than others in that they obtain persistently higher detection probability, the present data indicate that at least part of this enhancement of detection performance is associated with higher false alarm rates. Most of the variation among observers seems to be in the direction of varying threshold. If one observer were truly more sensitive than another, we would expect his  $P_D$  to be higher and his FAR to be lower. It is reasonable to conclude that there is not much variation in sensitivity among observers, but rather a greater degree of variation in decision threshold.

From the foregoing, we extract the following observations which will be used as guiding principles in the construction of our observer model:

- Observer sensitivity is primarily a function of background clutter.
- The intrinsic sensitivity of the observers is fairly uniform. Most of the variation between observers is in threshold.
- Threshold variations between observers are equivalent to threshold changes within an observer.

## V. MODEL DEVELOPMENT

Since the Phase I test data congregate so well along trajectories in the  $P_D$  vs. FAR plane, it seems reasonable to seek a description of the data in terms of signal detection theory. We demonstrate that the simplest such description is inadequate to fit the data, but that a slight generalization matches the data well.

We first must confront the problem that we have no *a priori* prescription for converting false alarm rates to probabilities. In order to know this relationship, we would have to know the number of times that each observer correctly rejected clutter objects—that is, the number of true dismissals. In a realistic search, we have no way to know the value of this quantity, nor even a clear idea of what it means. We therefore provisionally introduce a free normalization parameter, the false alarm opportunity rate (FAOR), to account for this unknown:

$$P_{FA} = \frac{\text{FAR}}{\text{FAOR}} .$$

The usual approach of SDT is to construct probability distribution functions (PDFs) related to the target and non-target ensembles. The simplest case is the one in which these two distributions are the same shape, but offset by some amount  $d$  which determines the sensitivity of the processor. The usual choice is a gaussian PDF, but we use a logistic PDF because the algebra is simpler and, in hindsight, it works well.

Our PDFs thus have the form  $\rho_T(x) = f(x)$  and  $\rho_{NT}(x) = f(x+d)$ , where

$$f(x) = \frac{e^x}{(1+e^x)^2} = \frac{1}{4} \operatorname{sech}^2 \frac{x}{2} .$$

The independent variable  $x$  may be interpreted as the logarithm of some signal strength parameter. A complete theory of human search performance would specify how to compute it. We are not in a position to do so, but are only using it as a vehicle to formally relate detection probability to false alarm rate. This is done by observing that

$$P_D(T) = \int_T^{\infty} \rho_T(x) dx ,$$

and

$$P_{FA}(T) = \frac{FAR(T)}{FAOR} = \int_T^{\infty} p_{NT}(x)dx ,$$

which we can solve by eliminating the threshold variable  $T$ . Observing that

$$F(x) \equiv \int_x^{\infty} f(x')dx' = \frac{1}{1+e^x} = \frac{1}{2} \left( 1 - \tanh \frac{x}{2} \right) ,$$

some algebra determines that the result for this case is

$$\frac{1}{P_D} = m \frac{1}{FAR} + b ,$$

where  $b = 1 - e^{-d}$  and  $m = FAOR e^{-d}$ .

Thus if we re-plot the data using the inverses of the  $P_D$  and FAR values, the points should fall along a straight line. Moreover the slope and intercept of the line should determine the two model parameters. Note that since  $d > 0$ , the intercept  $b$  must be on the interval  $(0,1)$ .

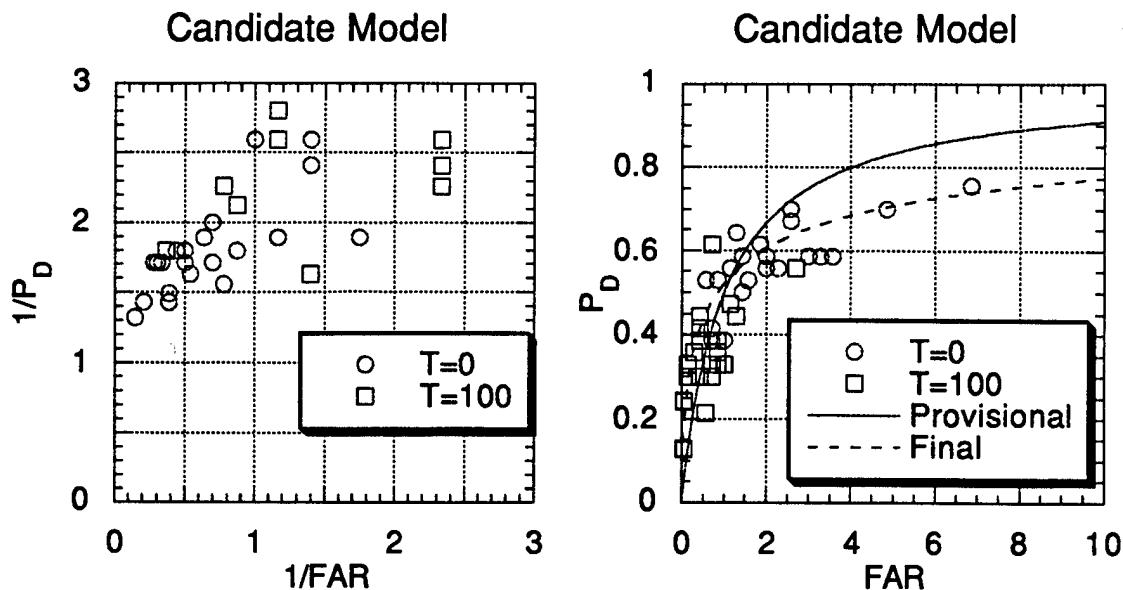
Unfortunately, plotting the data in the manner prescribed shows that the data are inconsistent with this simple model. Figure V-1, for view 5 from Phase I test b1, shows a case in point. As the left hand panel shows, the problem is that the intercept is clearly greater than one. If we display a representative model plot in the original coordinates, as in the right hand panel of Fig. V-1 (again from Phase I test b1 view 5), the problem appears to be that the theoretical curve does not saturate as fast as the data does at large FAR.

We speculate that the quality of the fit can be improved by relaxing the requirement that the two PDFs have the same shape. We therefore allow the non-target density to be narrower than the target density by a factor  $c$ . We then have

$$\text{logodds}(P_D) \equiv \ln\left(\frac{P_D}{1-P_D}\right) = \frac{1}{c} \ln\left(\frac{FAR}{FAOR-FAR}\right) - \frac{d}{c} .$$

This looks terrible; there are now three parameters instead of two and we still do not know how to compute the logodds for false alarms. But suppose we assert that we are in the low false alarm regime; in other words, that  $FAR \ll FAOR$ . This seems to be justified on intuitive grounds. It is not plausible that even the most prolific false alarm generator designates more clutter objects than he rejects. If we accept this, then a nice thing happens:

$$\ln\left(\frac{P_D}{1-P_D}\right) \approx \frac{1}{c} \ln(FAR) + \left( \frac{d}{c} - \frac{1}{c} \ln(FAOR) \right) .$$



**Figure V-1. Detection and false alarm data from Phase I test b1 view 5. The panel on the left shows a trend which intercepts the vertical axis at about 1.2. In the right panel the provisional model overshoots the high FAR points, but the final model accurately represents the trend.**

Now the relevant transform of the FAR is a simple logarithm and is independent of unknown parameters. Further, the separation parameter  $d$  has become inextricably entangled with the FAR normalization constant. We cannot separate them—but we do not have to. Just redefine an effective separation parameter which absorbs FAOR; call it  $\text{FAR}_{50}$ :

$$\text{FAR}_{50} = \text{FAOR} \exp(-d) .$$

Then we have

$$\ln(\text{FAR}) = c \ln\left(\frac{P_D}{1-P_D}\right) + \ln(\text{FAR}_{50}) .$$

As the name implies,  $\text{FAR}_{50}$  is the false alarm rate that is expected if the threshold is set so that  $P_D = 50$  percent. We now have a two-parameter description of the test data that (see Section VI) describes the observer response data quite well.

An unfortunate consequence of the need to go to a two-parameter description of false alarm phenomenology is that there is no single parameter that defines relative sensitivity. As Figure V-2 shows, if two curves have different values of the slope parameter  $c$ , then they will intersect. One of the curves will represent better performance on the low threshold (i.e., high  $P_D$ , FAR) side of the crossover point; on the high threshold side, the other will be better.

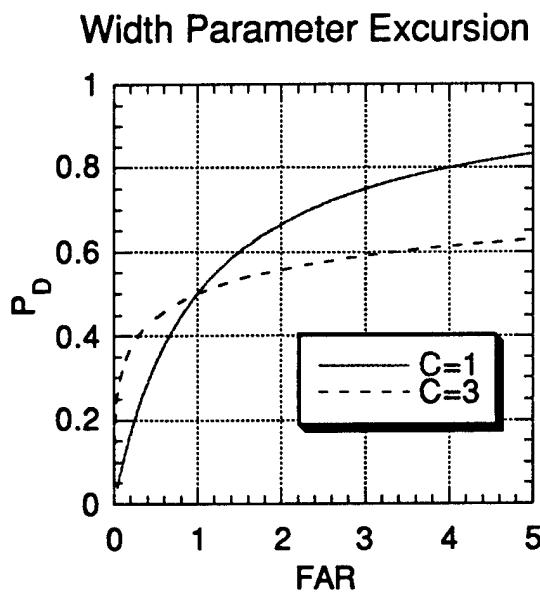


Figure V-2. The  $C=1$  curve shows greater sensitivity at  $FAR>1$ , but the  $C=3$  curve is better for  $FAR<1$ .

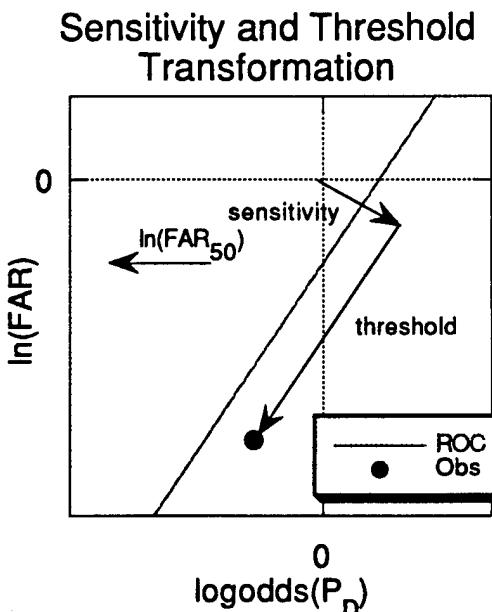


Figure V-3. Illustration of the definition of the threshold and sensitivity variables.

For a given trajectory, though, it is a simple matter to define a relative sensitivity, as well as a relative threshold value, for a given operating point. We illustrate the method in Figure V-3. The straight line is our fit to the data; the point represents the actual performance of an individual observer. For a given point, the sensitivity is the component of the displacement that is perpendicular to the line, the threshold is the component parallel to the line. The directions of the arrows in the figure give the positive sense of these quantities. Thus, for the point shown, the sensitivity and the threshold are both positive.

This description makes it clear that the transformation from the original coordinates to the sensitivity and threshold basis is a simple rotation, with the rotation angle specified by the slope of the characteristic line. The linear transformation to the new coordinates is given by the matrix formula

$$\begin{pmatrix} \text{threshold} \\ \text{sensitivity} \end{pmatrix} = \frac{1}{\sqrt{c^2+1}} \begin{pmatrix} -1 & -c \\ c & -1 \end{pmatrix} \begin{pmatrix} \text{logodds}(P_D) \\ \ln(FAR) \end{pmatrix}.$$

We emphasize that the sensitivity and threshold as defined here are constructed with respect to a specified slope parameter, given by  $c$ . Direct comparison of the individual values of these parameters between different background views with differing values of  $c$  is not meaningful. However, *differences* of values *within* a view—in particular, standard deviations of population statistics—are meaningful between views, at least to the extent that

our assumptions about the forms of the underlying PDFs are valid. It is the offset, not the scale, of the numbers that is arbitrary.

## VI. MODEL FITS AND POPULATION STATISTICS

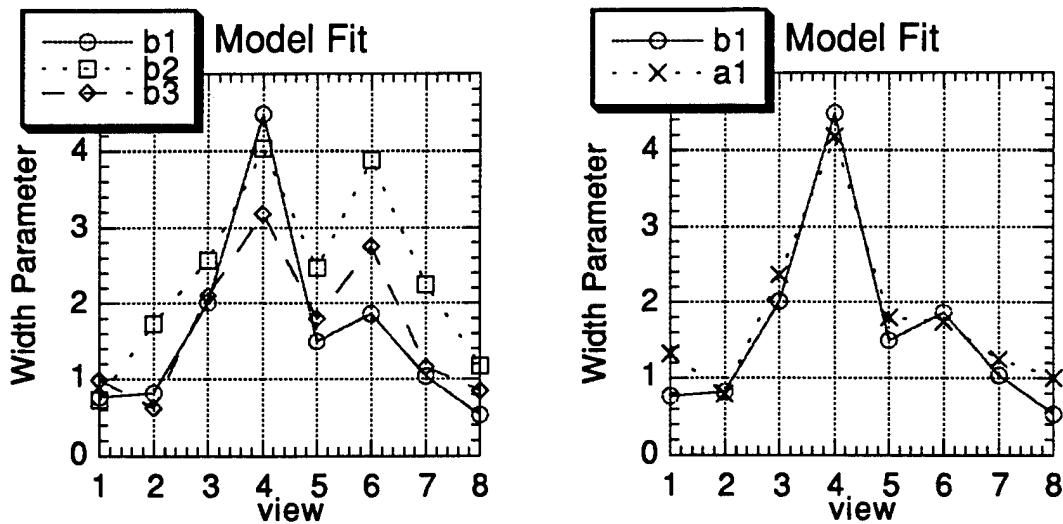
In this section we discuss the quality of the fits to the Phase I test data and consider general properties of the fitted parameters and population statistics. We defer the important question of predicting performance based on image statistics to Section VII.

Two parameters,  $\text{FAR}_{50}$  and the width parameter  $c$ , determine a characteristic curve for each view of each test. The determination of these parameters proceeds as follows: We first transform the  $P_D$ , FAR data to the  $[\text{logodds}(P_D), \ln(\text{FAR})]$  representation. Then, we determine the parameters of the characteristic curves (which are linear in the new basis) for the eight views of the four Phase I tests. The "fitting" procedure is simply to construct a straight line for each view of each test. Each line is determined by two points: the centroid (over observers) of the  $T = 100$   $[\text{logodds}(P_D), \ln(\text{FAR})]$  points, and the centroid of the  $T = 0$  points.

The displays of the complete set of 32 fits, along with a table containing the parameters of the fits, are displayed in Appendix A. Note that there is more scatter in the data corresponding to views 1, 2, 7, and 8 than the other four views. Because of the small number of false alarms in these views, the estimates of the logarithms of false alarm rates are subject to large statistical fluctuations. The situation is further exacerbated for views 1 and 2, since there are also few missed detections in those views.

Statistical fluctuations notwithstanding, our model for the data describes it quite well. At least for those views with adequate false alarm statistics (views 3 through 6), the data closely follow the lines that we have constructed.

In all views and all tests, the high confidence subset of the data extends smoothly as an extrapolation of the low confidence portion. Indeed, in every case there are some observers at the low threshold end of the high confidence group whose performance is essentially equivalent to some at the high threshold limit of the low confidence group. The individual observer's threshold thus seems to be under at least some degree of voluntary control. This suggests that the observer pool could be made more homogeneous through training or feedback; inter- and intra-observer threshold variations seem to be equivalent.

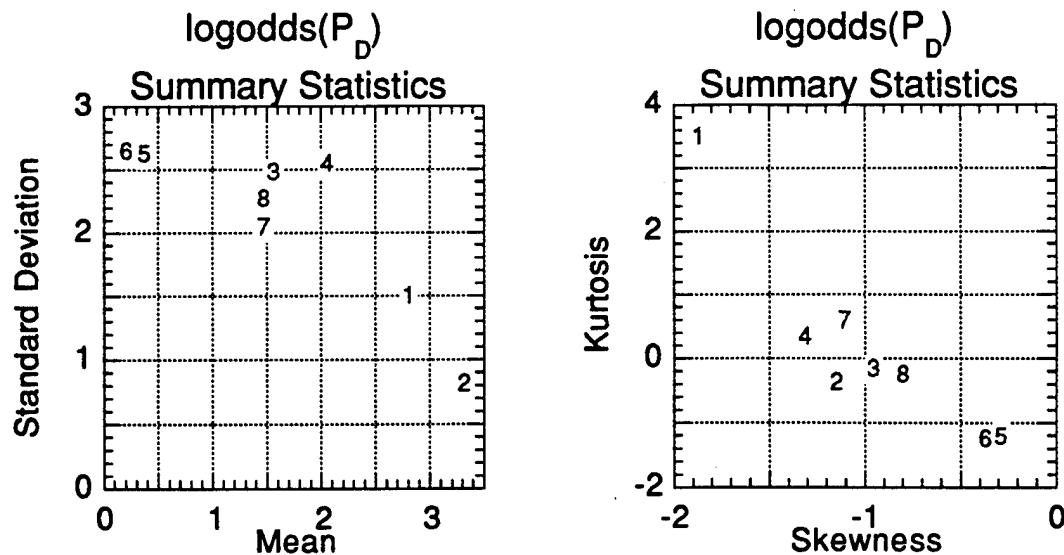


**Figure VI-1. Comparison of the dependence of the slope (or width) parameter,  $c$ , on view for the various sensor simulations (panel a, on left) and observer pools (panel b, on right).**

It is interesting that the slopes of the characteristic curves are quite different between views within a given test. It is perhaps even more interesting that the slopes are so similar between tests within a given view. Figure VI-1a displays the value of the slope parameter  $c$  for the b1, b2, and b3 tests. It shows that the trends from view to view persist for the different sensor simulations. Figure VI-1b shows an even more striking similarity between the a1 and b1 test, which represent different observer sets for the same sensor simulation.

While there is considerable correlation between views and the slope parameter, the correlation between the slope parameter and the image based clutter measures that we have considered is not sufficiently convincing to warrant elevating it to a model prescription. Instead we simply observe that the values of  $c$  cluster about a mean value of 2 with a standard deviation of 1.

It is worthwhile to try to understand the origin of the variability of the  $c$  parameter. Recall that the slope parameter arose as a description of the presumed underlying probability distribution function associated with the detectability of the targets which appear in the various views. Figure VI-2 shows a scatter plot of the summary statistics of the logodds of the detection probability associated with test b1. View 4 certainly does not appear to be anomalous in any sense that can be determined from this figure.

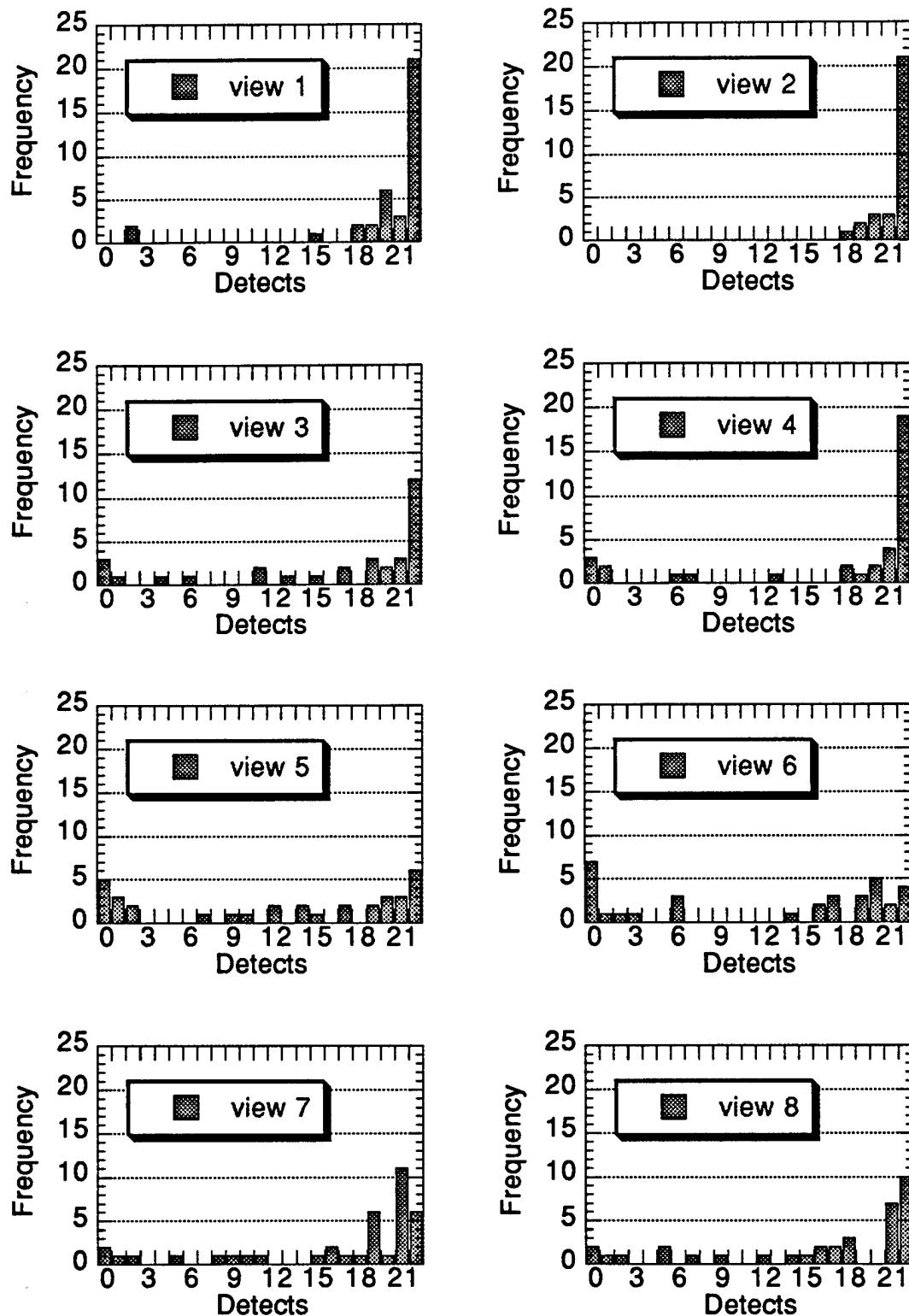


**Figure VI-2. Summary statistics of the  $\logodds(P_D)$  distribution for the various views. The points are labeled by the view number.**

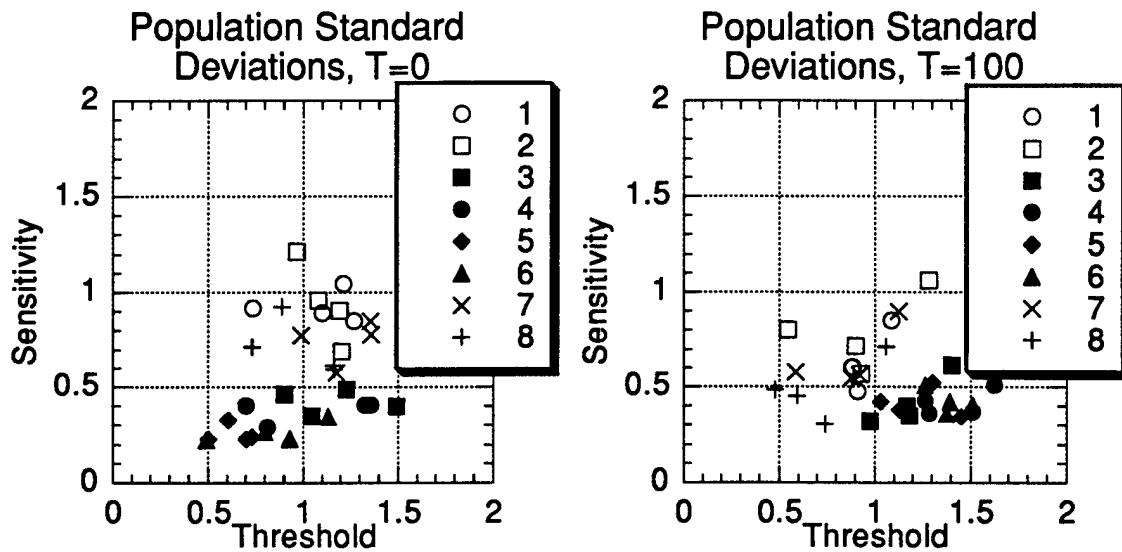
Nevertheless, we suspect that the variation in  $c$  has more to do with subtle aspects of the detectability of the targets in the test than with clutter *per se*. As evidence that this may be so, we present in Figure VI-3 a set of histograms, broken out by view, of the detection probability, at the  $T = 0$  confidence level, of the targets (that is, averaged over observers) as seen in the b1 test. Note that for view 4, which has the largest slope parameter ( $c = 4.4$ ), the histogram is somewhat anomalous with respect to the others: the target detection probabilities are almost all either very close to one or to zero. This means that there cannot be much variation in the observer detection probabilities (averaged over targets) due to the bimodal target set. In other words, the observers are constrained to a relatively constant detection probability, independent of threshold, while false alarms can vary freely. Thus, the large slope parameter.

Let us now turn to the set of summary statistics that describe the observer population.

The results for the population standard deviations are summarized in Figure VI-4. The two panels, a and b, correspond to the 0 percent and 100 percent confidence thresholds, respectively. The points corresponding to the views that have adequate false alarm statistics are denoted by solid symbols, while the less statistically significant results have crosses or open symbols.



**Figure VI-3. Frequency distribution of detections for the eight views of test b1. Note the relative lack of detections in the middle range of view 4.**

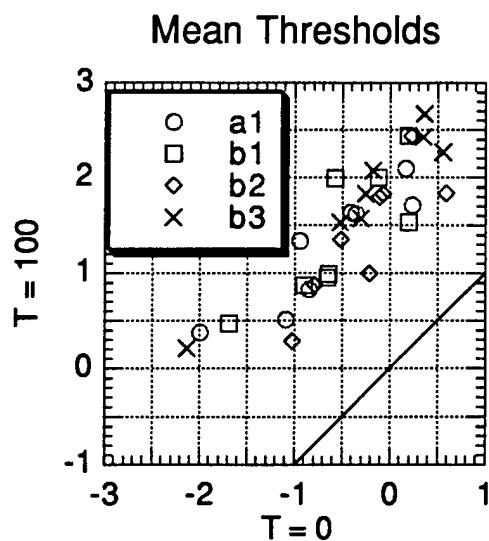


**Figure VI-4. Standard deviations of the observer populations for the 0% (left) and 100% confidence thresholds. The views with filled symbols are those with statistically significant estimates. The four points per view correspond to the four Phase I tests a1, b1, b2, and b3.**

The quantitative results shown here buttress the qualitative remarks that we have made previously: the population is much more homogeneous in sensitivity than in thresholds. In fact, for the statistically significant views, the thresholds standard deviations are a factor of three larger than the sensitivity standard deviations; this holds within both the low and high confidence observer declaration thresholds.

The other relevant quantity that can be derived is the average threshold offset between the high and low confidence declaration thresholds. As Figure VI-5 shows, this offset is just as persistent as the standard deviations. All of the points show an offset of very close to two units. However much the observers seem to disagree on the absolute definitions of 0 percent confidence and 100 percent confidence, they agree remarkably well on the relative difference between these two thresholds. Note also that the two-unit separation between the centroids of the 0 percent and 100 percent distributions is smaller than the sum of their threshold standard deviations; as noted previously, the two distributions overlap.

The model we have developed in the preceding section produces a well-defined set of persistent parameters that describe the observer population. These population parameters are summarized in Table VI-1.



**Figure VI-5. Thresholds for 100% confidence vs. 0% confidence for all views and tests, averaged over observers.**  
**The line corresponds to equal thresholds.**

**Table VI-1. Observer population statistics derived from Phase I Target Acquisition tests**

Test No.	View No.	Centroids of observers in (logodds $P_D$ , log FAR) basis				Slope, intercept of line joining centroids				Centroids of observers in (sensitivity, threshold) basis				Std. deviations of observers in (sensitivity, threshold) basis			
		T = 0		T = 100		T = 0		T = 100		T = 0		T = 100		T = 0		T = 100	
		dkp $\mu$	dkp $\sigma$	dhj $\mu$	dhj $\sigma$	t $\mu$	t $\sigma$	thj $\mu$	thj $\sigma$	thj $\mu$	thj $\sigma$	th $\mu$	th $\sigma$	th $\mu$	th $\sigma$	th $\mu$	th $\sigma$
a1	1	2.12	-1.09	0.887	-2.72	1.32	-3.89	-0.409	2.35	1.63	2.35	0.737	0.916	1.09	0.853		
a1	2	3.34	-1	1.49	-2.47	0.794	-3.65	-1.99	2.86	0.375	2.86	0.965	1.21	1.28	1.06		
a1	3	0.731	.0724	-0.037	-1.74	2.36	-1.65	-.352	0.645	1.62	0.645	0.903	0.462	1.4	0.615		
a1	4	1.12	0.705	0.585	-1.51	4.17	-3.95	-.946	0.921	1.33	0.921	0.702	0.403	1.63	0.509		
a1	5	0.126	0.897	-0.686	-0.564	1.8	0.669	-.846	-.325	0.827	-.325	0.504	0.225	1.45	0.344		
a1	6	0.252	1.11	-0.546	-0.272	1.74	0.676	-1.09	-.338	0.508	-.338	0.495	0.223	1.51	0.403		
a1	7	0.928	-0.962	-0.277	-0.246	1.24	-2.12	0.168	1.33	2.09	1.33	0.989	0.775	1.12	0.901		
a1	8	1.01	-1.35	-0.037	-2.39	0.994	-2.35	0.239	1.67	1.71	1.67	0.733	0.714	1.06	0.713		
b1	1	2.34	-1.99	1.07	-2.96	0.765	-3.78	-.651	3	0.954	3	1.1	0.891	0.885	0.604		
b1	2	3.55	-1.68	1.88	-3.04	0.816	-4.57	-1.69	3.54	0.469	3.54	1.19	0.908	0.899	0.716		
b1	3	1.04	-.389	0.099	-2.28	2	-2.47	-.117	1.1	1.99	1.1	1.23	0.489	1.17	0.4		
b1	4	1.3	0.301	0.744	-2.2	4.48	-5.53	-.578	1.21	1.99	1.21	0.813	0.287	1.51	0.367		
b1	5	0.287	0.575	-0.614	-.776	1.5	0.145	-.637	-.080	0.987	-.080	0.702	0.226	1.03	0.422		
b1	6	0.249	0.891	-.591	-0.671	1.86	0.428	-.903	-.203	0.871	-.203	0.799	0.263	1.26	0.507		
b1	7	1.16	-1.39	-0.4	-3	1.03	-2.59	0.194	1.8	2.44	1.8	1.35	0.849	0.921	0.562		
b1	8	0.998	-2.31	-0.17	-2.93	0.532	-2.84	0.204	2.51	1.53	2.51	0.888	0.924	0.739	0.304		

**Table VI-1 (continued)**

Test No.	View No.	Centroids of observers in (logodds $P_D$ , log FAR) basis				Slope, intercept of line joining centroids				Centroids of observers in (sensitivity, threshold) basis				Std. deviations of observers in (sensitivity, threshold) basis			
		T = 0		T = 100		C		B		T = 0		T = 100		T = 0		T = 100	
		dkμ	flμ	dhμ	fhμ	tlμ	thμ	slμ	shμ	tlσ	thσ	slσ	shσ	thσ	shσ	thσ	shσ
b2	1	2.83	-2.17	1.78	-2.93	0.726	-4.22	-1.02	3.42	0.285	3.42	1.21	1.05	0.878	0.602		
b2	2	3.87	-1.99	3.27	-3.04	1.73	-8.68	-216	4.35	0.996	4.35	1.2	0.687	0.544	0.802		
b2	3	1.2	-345	0.505	-2.12	2.57	-3.42	-112	1.24	1.79	1.24	1.04	0.348	1.19	0.346		
b2	4	1.66	-342	1.21	-2.18	4.03	-7.05	-0.68	1.7	1.83	1.7	1.36	0.405	1.26	0.427		
b2	5	0.701	0.567	.0745	-977	2.47	-1.16	-789	0.436	0.877	0.436	0.607	0.326	1.13	0.377		
b2	6	0.535	0.392	.0712	-1.41	3.89	-1.69	-513	0.42	1.35	0.42	0.929	0.229	1.37	0.356		
b2	7	1.38	-864	0.479	-2.87	2.24	-3.95	0.227	1.61	2.43	1.61	1.18	0.578	0.878	0.547		
b2	8	1.35	-1.93	0.547	-2.88	1.17	-3.52	0.593	2.29	1.84	2.29	1.16	0.594	0.594	0.452		
b3	1	2.05	-1.66	0.717	-2.97	0.983	-3.67	-304	2.62	1.57	2.62	1.27	0.85	0.91	0.476		
b3	2	3.67	-1.86	1.69	-3.1	0.624	-4.15	-213	3.52	0.206	3.52	1.08	0.96	0.931	0.571		
b3	3	0.777	-753	-117	-2.63	2.1	-2.39	0.346	1.03	2.43	1.03	1.49	0.399	0.975	0.322		
b3	4	1.28	-211	0.605	-2.36	3.18	-4.29	-182	1.28	2.07	1.28	1.33	0.407	1.28	0.36		
b3	5	0.149	0.513	-844	-1.27	1.8	0.245	-521	-119	1.52	-119	0.732	0.239	1.3	0.521		
b3	6	.0605	0.248	-0.65	-1.71	2.75	.0815	-253	-0.028	1.82	-0.028	1.13	0.342	1.39	0.422		
b3	7	1.04	-1.38	-466	-3.12	1.16	-2.58	0.361	1.69	2.66	1.69	1.36	0.78	0.588	0.576		
b3	8	0.902	-1.95	-404	-3.04	0.839	-2.7	0.559	2.07	2.26	2.07	1.16	0.615	0.481	0.483		

## VII. FALSE ALARM PREDICTION FROM IMAGE STATISTICS

The statistical variance (SV) statistic has been studied extensively as a predictor of target detection probability.<sup>5,6</sup> In this section we connect it directly to false alarm rate.

SV is simply the square root of the average of the local variance of the image pixel values. That is,  $S = \sqrt{<[(1-F) \cdot I]^2>}$ , where  $I$  is the original image,  $1$  is the unit impulse filter, and  $F$  is a boxcar filter 23 pixels square. (See source in footnote 6 for discussion of this choice.) The symbol  $\cdot$  denotes convolution, and  $<>$  the average over all pixels.

The choice of this statistic as the one to connect to false alarm performance is based upon two considerations. First, it is already in use in the NVESD target detection model. Second, it is defined unambiguously, up to a scale factor that defines the size of the region over which the variances are computed. Thus it allows little "tuning" or calibration.

We mention here our earlier work, heretofore unpublished, in which we attempted to use the SV density to predict *specific* false alarm locations. This work had limited success. While the false alarms often cluster about such peaks, a substantial fraction of the time they do not. Furthermore, there are always some other peaks with equally high or higher SV value that are not correlated with false alarms. For a synopsis of our analysis of SV as a deterministic predictor of false alarms, see Appendix B.

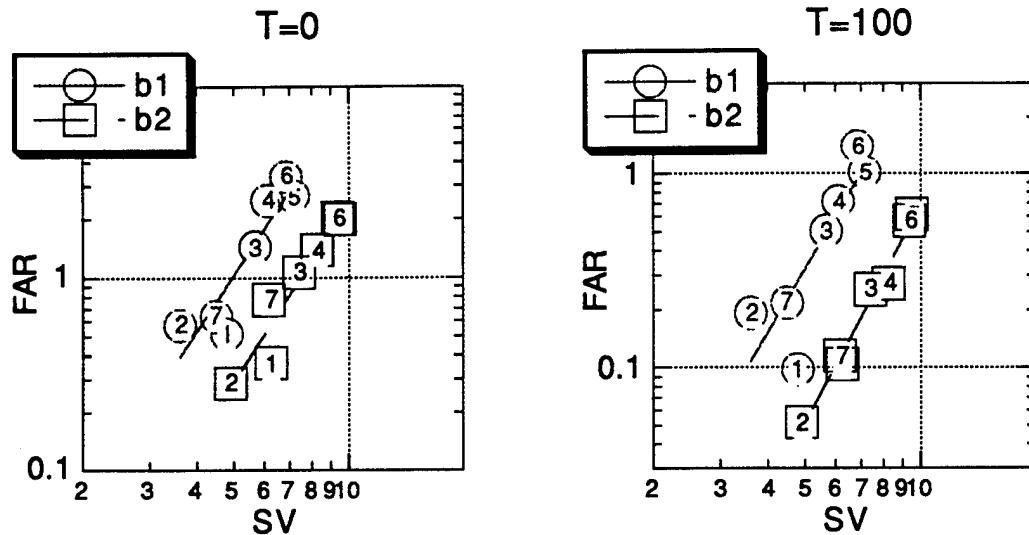
Despite the unreliability of SV as a tool in locating individual false alarm attractors, the positive but sporadic correlation with false alarm clusters gives us confidence that by averaging over a large region, such as an entire field of view, the sporadic component might be diminished. The hope is that for a given degree of complexity, as measured by SV, a more or less fixed fraction will be unidentifiable and therefore a false alarm candidate.

---

<sup>5</sup> D.E. Schmieder, M.R. Weathersby, *Detection Performance in Clutter with Variable Resolution*, IEEE Transactions on Aerospace and Electronic Systems, AES-19 #4, 1983.

<sup>6</sup> James D. Silk, *Statistical Variance Analysis of Clutter Scenes and Application to a Target Acquisition Test*, Institute for Defense Analyses, IDA Paper P-2950, 1994.

With this preamble in mind, we consider the test data. Figure VII-1 displays the correlation between false alarm rate and SV value. The quality of the correlation is quite satisfying. We are further encouraged that the logarithmic plots of the data are all linear, and all have roughly the same slope. It therefore seems appropriate to constrain the fit to be a power law.



**Figure VII-1. Correlation of false alarm rate with statistical variance, for low (left) and high (right) confidence detection. The statistical variance computation referenced here is SV<sub>23</sub> of footnote 6, in gray scale units.**

The results of power law fits to the data are displayed in the figure and tabulated in Table VII-1. The sensor and threshold dependent coefficients should not be regarded as being predictive. They may be useful for quantifying differences between sensors under controlled conditions, for example. But as the discussion of Section III demonstrates, the degree of sensitivity of false alarm results to observer state and expectations render any absolute estimate of false alarm rate based on system or image parameters extremely tenuous, at best. We believe that the best use of this model would be to predict excursions from a measured baseline performance.

Having made some estimate of false alarm rate based on image statistics, and an estimate of the ensemble averaged detection probabilities based on the NVESD detection model, it is now possible to generate operating characteristics of the form defined in Section V.

**Table VII-1. Results of power law fits of the form  
 $FAR = \text{constant} \times SV^{\text{power}}$  to the data of Figure VII-1.**

Confidence	Sensor	Constant	Power
T = 0	b1	0.0088	2.97
T = 0	b2	0.0021	3.06
T = 100	b1	0.0015	3.35
T = 100	b2	0.00013	3.72

The only remaining connection to be made is that while estimates of the average  $P_D$  can be provided by the Thermal TAMIP detection model, and the average FAR can be obtained from Table VII-1, the model of Section V requires the means of different variables, namely sensitivity and threshold. The final equation of Section V is the connection between these sets of variables. but is not strictly applicable to the mean values of these quantities, since the relation is nonlinear. It can be used to get a first order estimate. An improved estimate is obtained by expanding the observer probability distribution function, which is presumed to be bivariate gaussian in the sensitivity and threshold coordinates, to second order about the mean. Appendix C provides a Mathcad™ script which illustrates the procedure in detail.

## VIII. SUMMARY

We have presented a model of false alarms that is consistent with the NVESD Phase I test data. The model may loosely be partitioned into descriptive and predictive components. The descriptive part of the model, which prescribes how to treat the observer ensemble for a given average level of detection probability and false alarm rate, is a realistic representation of actual human behavior and is extensible to more general situations. The predictive portion, which related false alarm rate to image parameters, is demonstrably too brittle to extend beyond the specific conditions that were replicated in the Phase I tests.

We conclude by summarizing the algorithm for generating simulated observer responses. End-to-end implementation of the model described and justified in this work proceeds as follows:

- Specify the sensor, decision confidence level, and the background and target characteristics.
- Estimate the SV from the background and sensor based on image data if available.
- Determine the average expected FAR from sensor, decision confidence, and SV based on Section VII power law estimates.
- Determine the average  $P_D$  from sensor, decision confidence, target characteristics, and SV based on the Thermal TAMIP detection model.
- Select the slope parameter  $c$  based on the advice of Section VI (range 1 to 4 with mean of 2).
- Compute mean sensitivity and mean threshold from the mean  $P_D$ , mean FAR, and  $c$  using the formula at the end of Section V. (This is only an approximation, but a fairly good one, for the mean values; a better but more complicated one is provided in Appendix B.)
- Set observer population standard deviations for the sensitivity and thresholds based on the recommendations of Section VI (see especially Figure VI-4; for full detection with  $T = 100$ , we recommend 0.4 and 1.3 for sensitivity and threshold, respectively).
- Draw individual observer sensitivity and threshold from these distributions.
- Convert to individual  $P_D$  and FAR based on inverse of the formula at the end of Section V.

**APPENDIX A**

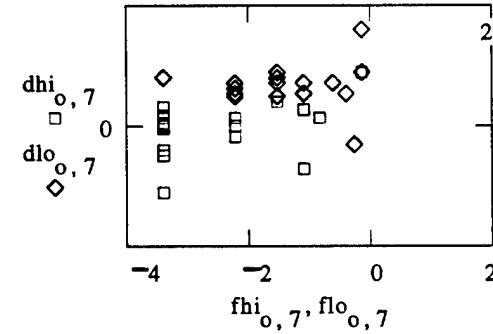
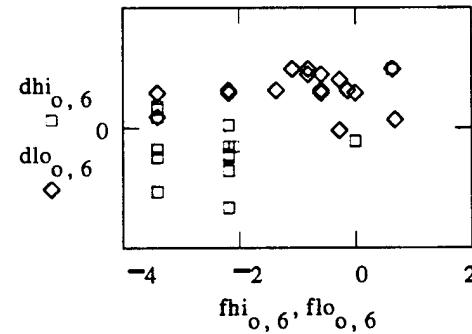
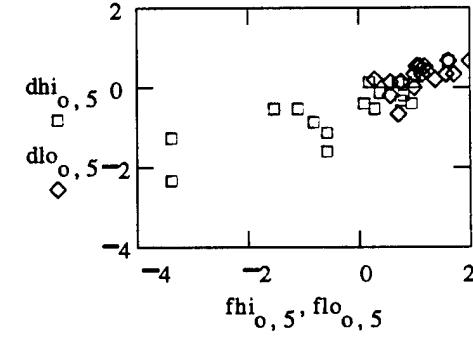
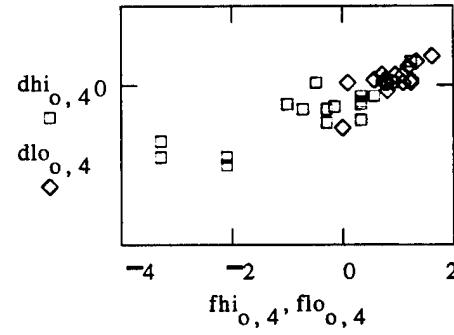
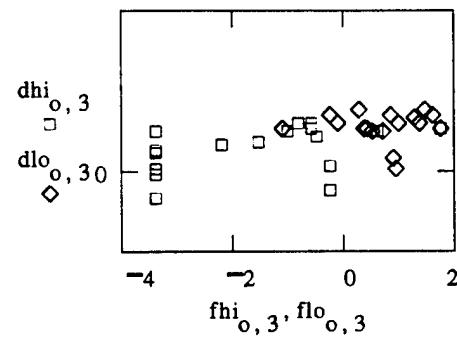
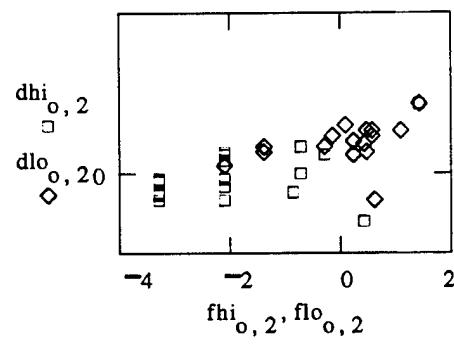
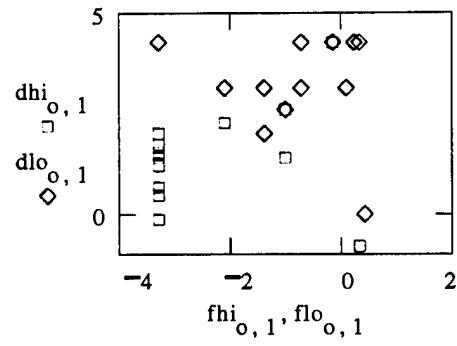
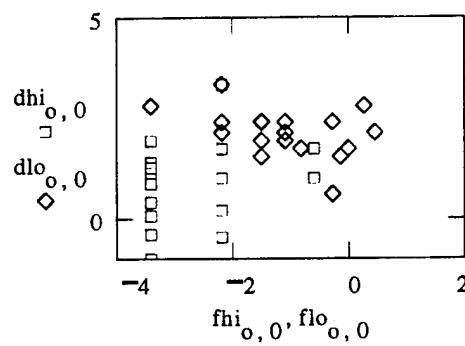
**OBSERVER TEST DATA**

## **APPENDIX A**

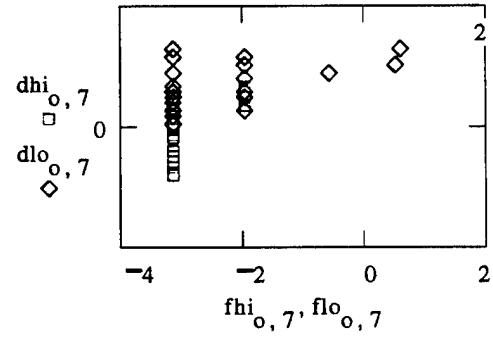
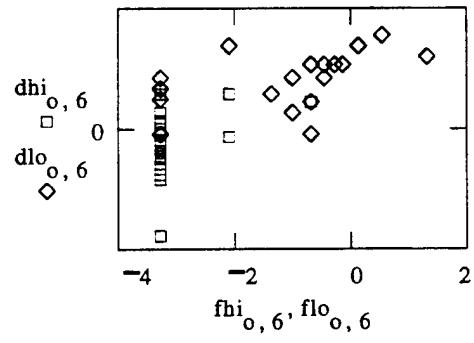
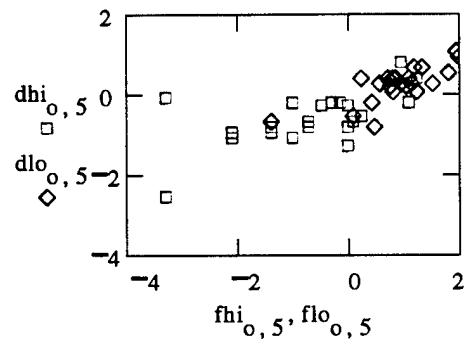
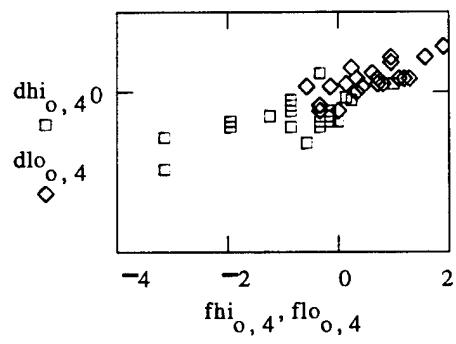
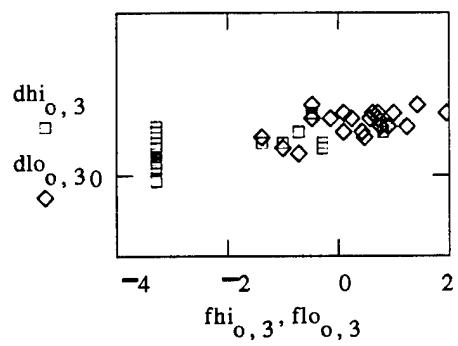
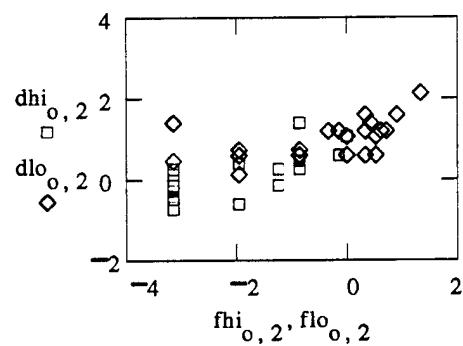
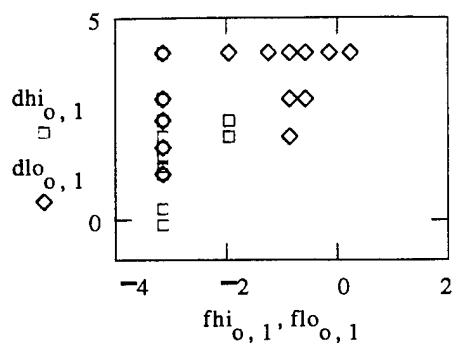
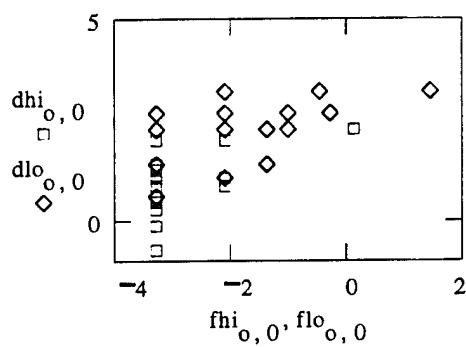
### **OBSERVER TEST DATA**

The following pages contain plots of the complete set of observer data from the NVESD Phase 1 test. On each page the test index (0, 1, 2, 3) denotes the NVESD Phase I test (a1, b1, b2, b3).

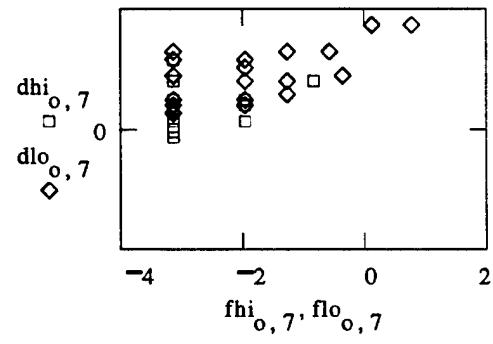
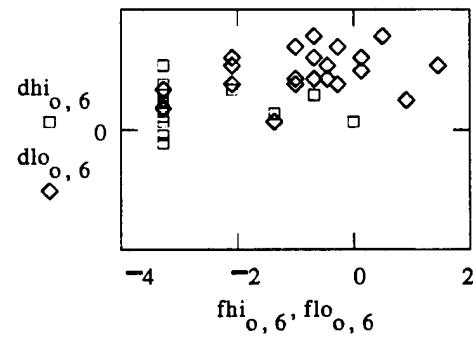
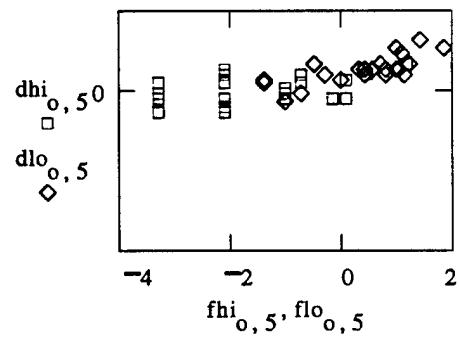
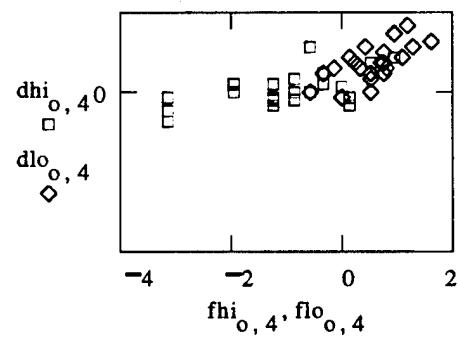
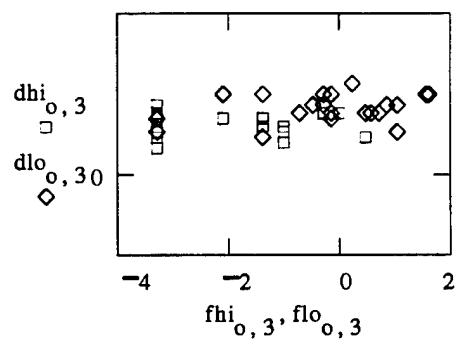
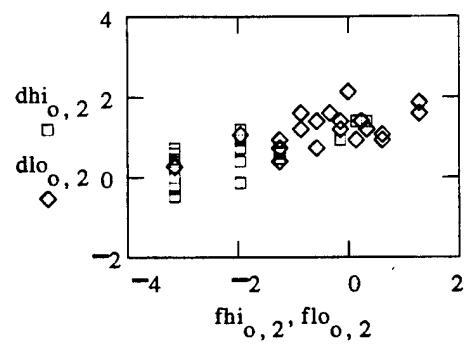
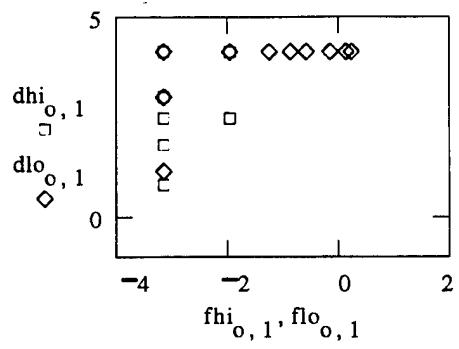
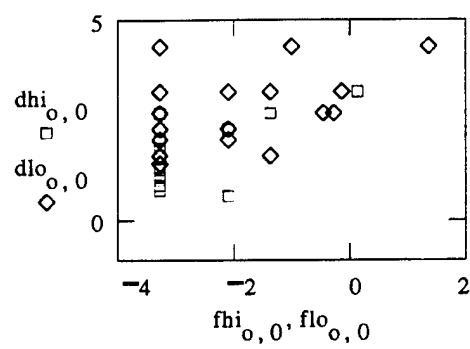
it = 0



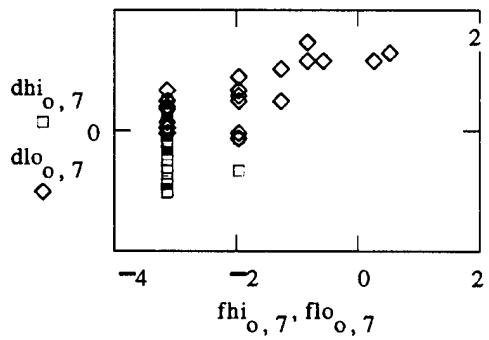
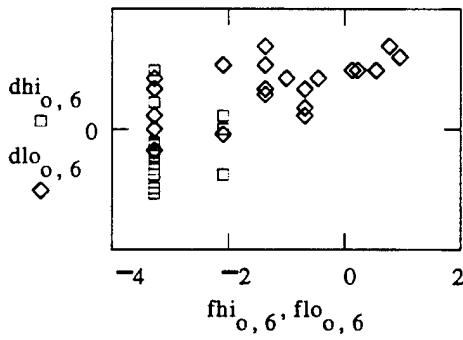
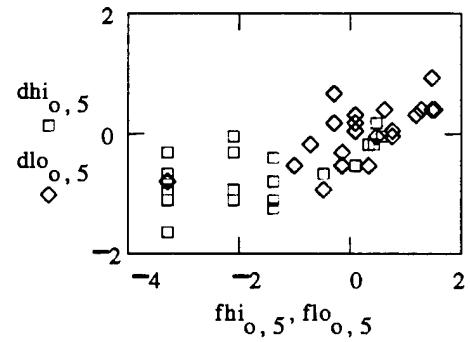
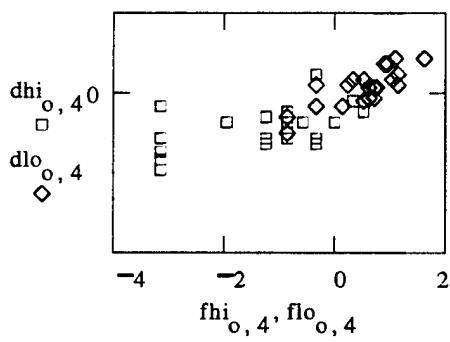
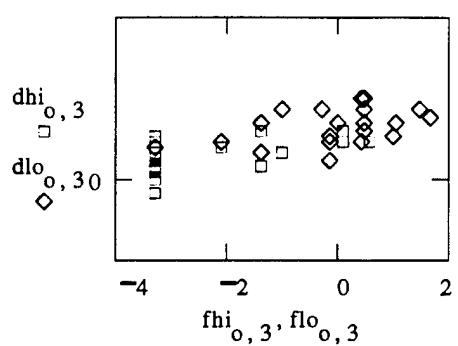
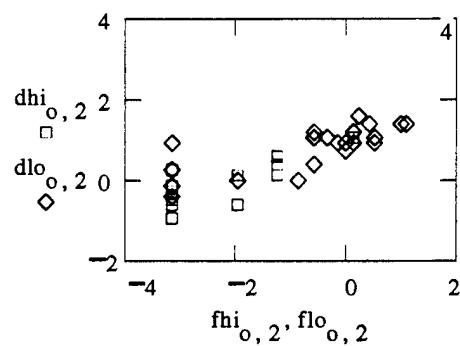
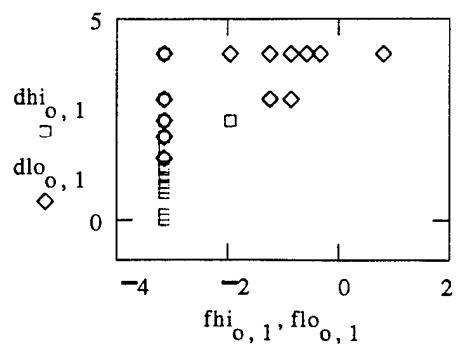
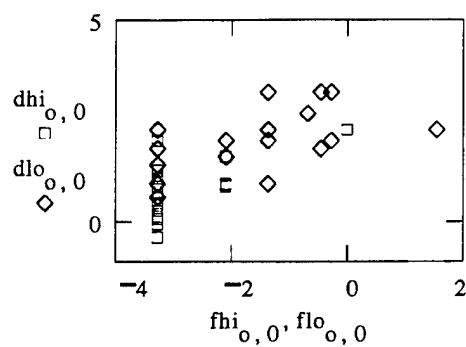
it = 1



it = 2



it = 3



**APPENDIX B**

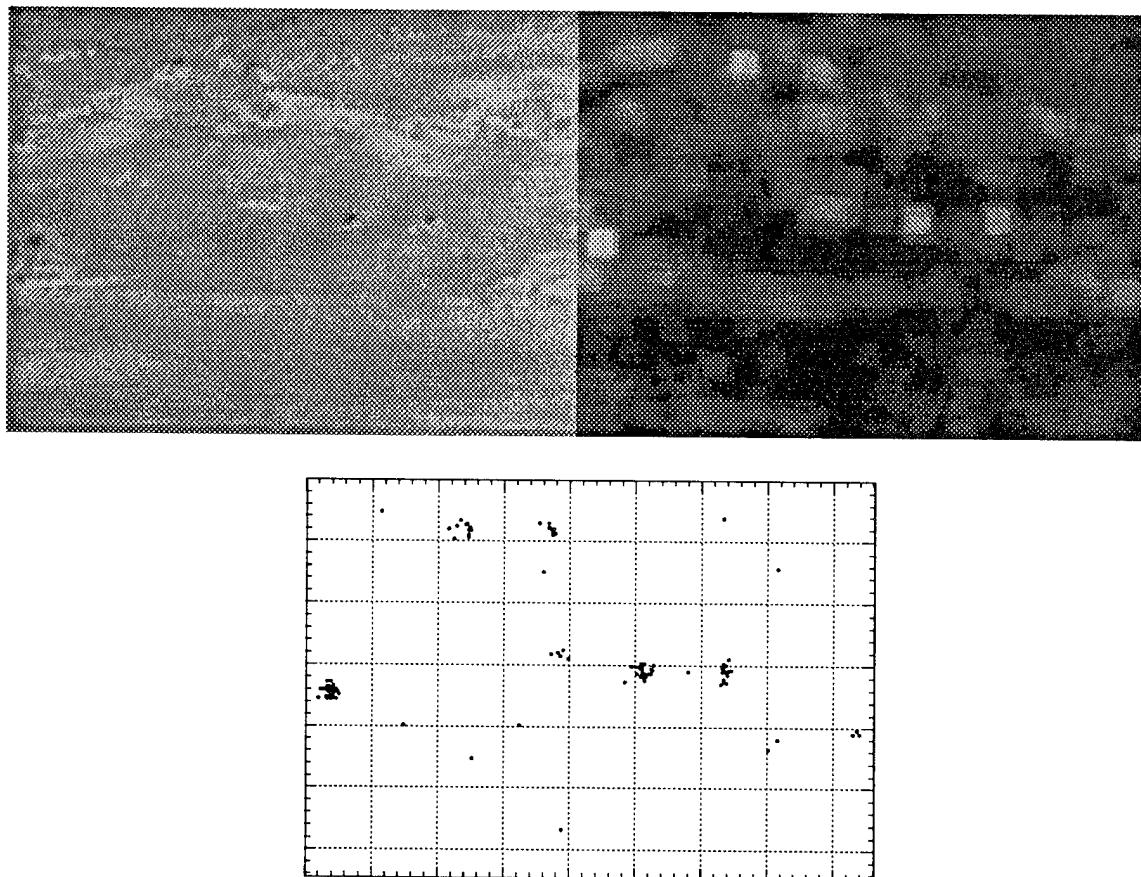
**FALSE ALARM LOCATION ANALYSIS**

## APPENDIX B

### FALSE ALARM LOCATION ANALYSIS

In Section VII we alluded to an early effort to establish the statistical variance (SV) statistic as a predictor of false alarm locations. Since that analysis has not been discussed elsewhere, we document it here for the record.

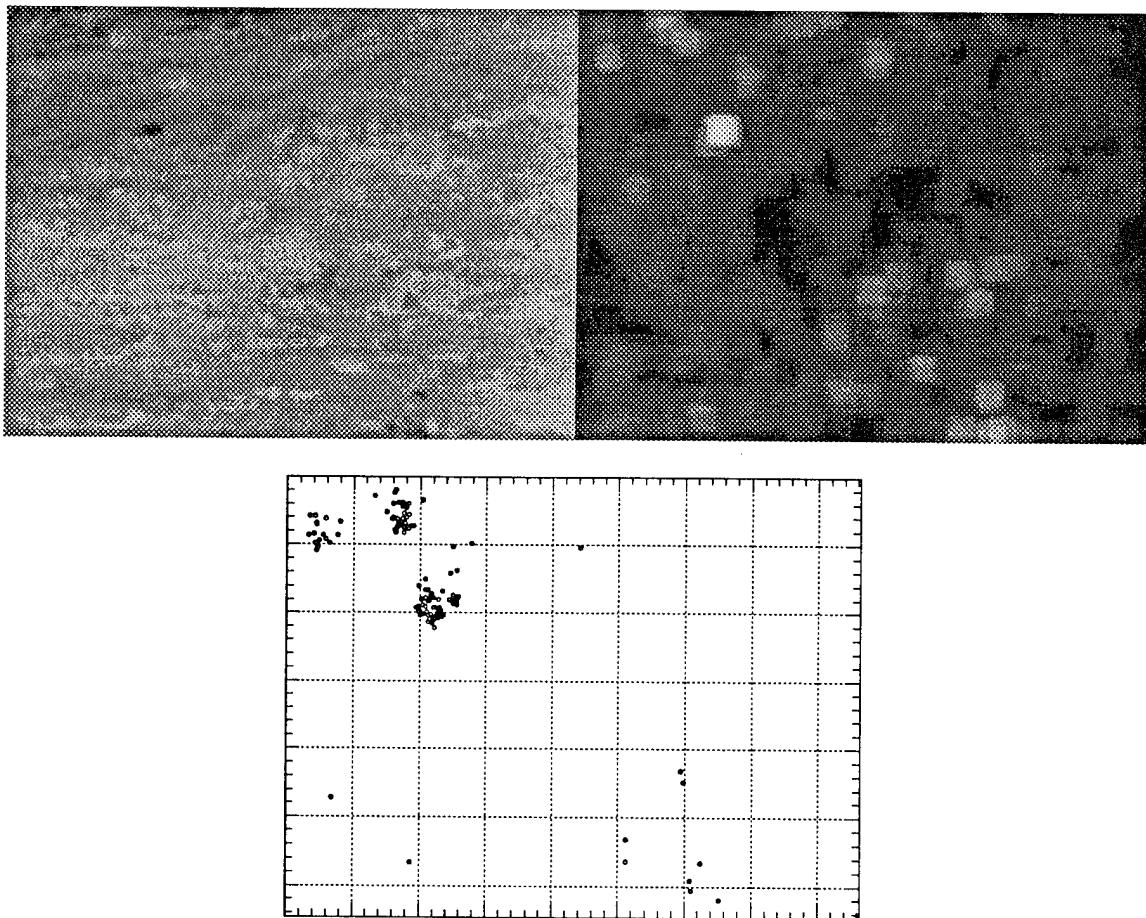
The first step in the analysis was to subjectively evaluate the correlation between the SV density image and the declared false alarms. The meaning and method of construction of the SV image are presented in the source in footnote 6.



**Figure B-1.** Top left: Scene 2225. Top right: Corresponding SV density image. Bottom: False alarm locations.

Two cases are shown here. In Figure B-1 is the comparison for background view 2225. The qualitative correlation seems strikingly good; the inverted "big dipper" pattern is evident in all three pictures.

Unfortunately, view 2225 looks considerably better than the others. View 2741, shown in Figure B-2, is more typical. While the strongest SV peak is also the strongest false alarm attractor, the correlation over the rest of the image seems weak.



**Figure B-2. As Figure B-1, but background 2741.**

Furthermore, the quantitative correlation between SV value and local false alarm density is not as strong as one would like. Our method of analysis is as follows: Since the clusters seem to correlate with SV peaks, we first need to construct a clustering algorithm to "score" the false alarm clusters (recall there is no "ground truth" for the clutter). We analyzed all clusters of two or more false alarms, except for view 2747, where there were so many small clusters that we stopped at the top 11. Then a rather similar algorithm was

used to locate and evaluate SV peaks. We collected the top 20 SV peaks for each background view.

Finally, we correlated the false alarm clusters with the SV peaks. For each false alarm cluster, the distance was computed to all the SV peaks on the top 20 list. The closest SV peak on the list was associated with the given cluster, provided that it was within 20 pixels—about the size of the average target. If no peak was within that range, no association was made.

The results of this analysis are shown in Table B-1. Observe that view 4 shows much better correlation than any of the other views. Overall, fewer than 60 percent of the clusters are associated with SV peaks, even though the top 20 peaks were identified for each view to accommodate an average of seven clusters per view. We do not believe that this level of correlation between SV and false alarm locations justifies the formulation of a deterministic model of false alarms. We have therefore confined ourselves to the average treatment of Section VII.

**Table B-1. Correlation of false alarm locations with SV value.**

FA	SV	view	FA	SV	view
	seq			seq	
3	2	11.1	1	1701	52
2	—	—	1	1701	36
8	4	9.3	2	1705	22
2	—	—	2	1705	26
2	—	—	2	1705	17
2	3	9.4	2	1705	10
2	12	6.9	2	1705	9
21	9	12.5	3	2221	7
12	13	11.6	3	2221	5
17	1	18.0	3	2221	5
5	—	—	3	2221	5
2	—	—	3	2221	5
2	—	—	3	2221	6
32	1	19.0	4	2225	3
23	7	14.3	4	2225	2
15	8	14.2	4	2225	2
11	2	17.4	4	2225	2
8	6	14.9	4	2225	2
5	15	12.4	4	2225	12
3	5	15.0	4	2225	3
2	—	—	4	2225	2

(Continued)

Table B-1 (Continued)

FA	SV	view	FA	SV	view				
FA	seq	SV	#	View	FA	seq	SV	#	View
67	1	26.9	5	2741		2	—	8	3265
31	9	13.9	5	2741					
17	17	12.6	5	2741					
3	—		5	2741					
4	—		5	2741					
3	5	15.3	5	2741					
2			5	2741					
2	8	14.5	5	2741					
2	14	12.8	5	2741					
2	20	12.3	5	2741					

We believe that the reason for the sporadic effectiveness of SV as a predictor of specific false alarm objects is due to the presence of higher cognitive processes in target detection. Some objects are simply identified by the observer as what they are—a bush or rock, for instance—and are excluded from being declared as a detection on this basis. Although the detection process is often hypothesized to be a lower level process than the recognition and identification tasks, these higher elements seem to be present in military detection, whether we like it or not.

## APPENDIX C

### TRANSFORMATION OF AVERAGED VARIABLES

## APPENDIX C

### TRANSFORMATION OF AVERAGED VARIABLES

In order to simulate the detection and false alarm performance of an ensemble of observers, we need to know five parameters:  $\bar{s}$  and  $\bar{t}$ , the means of sensitivity and threshold;  $\sigma_s$  and  $\sigma_t$ , their standard deviations; and  $c$ , the slope parameter. In the foregoing work we have recommended standard choices for the latter three quantities. In this appendix we explicitly show how to determine  $\bar{s}$  and  $\bar{t}$  from the quantities that are predicted, namely  $\overline{\text{FAR}}$  and  $\overline{P_D}$ .

We start by writing our observables as functions of sensitivity and threshold:

$$P_D(s, t) = \frac{1}{1 + \exp(-as + bt)}, \quad \text{FAR}(s, t) = \exp(-bs - at) .$$

We assume that the observer ensemble is described by a Gaussian probability distribution function:

$$\phi(s, t) = \frac{1}{2\pi\sigma_s\sigma_t} \exp\left[-\frac{(s - \bar{s})^2}{2\sigma_s^2} - \frac{(t - \bar{t})^2}{2\sigma_t^2}\right] .$$

Thus the averages of the observable quantities over the observer ensemble are given by

$$\overline{P_D} = \int P_D(s, t) \phi(s, t) ds dt, \quad \text{and} \quad \overline{\text{FAR}} = \int \text{FAR}(s, t) \phi(s, t) ds dt .$$

These integrals are hard, but it is enough to expand the functions  $P_D$  and  $\text{FAR}$  to second order about the mean values  $\bar{s}$  and  $\bar{t}$ . The first order terms cancel because of the symmetry of  $\phi$ , so we get

$$\overline{P_D} = [1 + 2A(1 - P_D(\bar{s}, \bar{t}))^2] P_D(\bar{s}, \bar{t}) \quad \text{and} \quad \overline{\text{FAR}} = [1 + B] \text{FAR}(\bar{s}, \bar{t}) ,$$

where

$$A = \frac{\sigma_t^2 - c^2 \sigma_s^2}{c^2 + 1} \quad \text{and} \quad B = \frac{\sigma_s^2 + c^2 \sigma_t^2}{c^2 + 1} .$$

The expressions for  $\overline{P_D}$  and  $\overline{\text{FAR}}$  can now be inverted; the first is cubic in  $P_D$ , but an approximate iterative solution suffices. The final step is to use the definitions of  $P_D$  and  $\text{FAR}$  to solve for  $\bar{s}$  and  $\bar{t}$ . The following Mathcad™ script generates pseudodata and illustrates these transformations.

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

<b>1. AGENCY USE ONLY (Leave blank)</b>			<b>2. REPORT DATE</b> September 1995	<b>3. REPORT TYPE AND DATES COVERED</b> Final—June 1994 - June 1995
<b>4. TITLE AND SUBTITLE</b> A Model of False Alarms in Target Acquisition by Human Observers			<b>5. FUNDING NUMBERS</b> DASW01 94 C 0054 A-162	
<b>6. AUTHOR(S)</b> James D. Silk				
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Institute for Defense Analyses 1801 N. Beauregard St. Alexandria, VA 22311-1772			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b> IDA Paper P-3076	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> Advanced Research Projects Agency 3701 N. Fairfax Drive Arlington, VA 22203-1714			<b>10. SPONSORING/MONITORING AGENCY REPORT NUMBER</b>	
<b>11. SUPPLEMENTARY NOTES</b>				
<b>12a. DISTRIBUTION/AVAILABILITY STATEMENT</b> Approved for public release; distribution unlimited.			<b>12b. DISTRIBUTION CODE</b>	
<b>13. ABSTRACT (Maximum 180 words)</b> The modeling of target acquisition by human observers has focused on the problem of predicting whether targets will be detected. The closely related issue of false alarm prediction has received less attention. While predicting false alarms is secondary to true detection, it is nevertheless important to understand the effects of false alarms and to account for them in the development of doctrine. In this work we extend the scope of target acquisition modeling to the consideration of false detections. The model is based on the analysis of data obtained in a series of target acquisition tests. It is phenomenological in the sense that it seeks only to describe the results of the tests. An important finding from the analysis of the test data is that the dominant determinant of false alarm rate is the expectation of the human subject. A more general review of the test results reveals features that strongly suggest a description based on signal detection theory. Re-analysis of the test data in this context allows us to construct such a description and to extract the parameters that describe the observer ensemble. Finally, we demonstrate the correlation between the mean false alarm rate and a scene complexity statistic.				
<b>14. SUBJECT TERMS</b> model, target acquisition, false alarms			<b>15. NUMBER OF PAGES</b> 50	
<b>16. PRICE CODE</b>				
<b>17. SECURITY CLASSIFICATION OF REPORT</b> UNCLASSIFIED	<b>18. SECURITY CLASSIFICATION OF THIS PAGE</b> UNCLASSIFIED	<b>19. SECURITY CLASSIFICATION OF ABSTRACT</b> UNCLASSIFIED	<b>20. LIMITATION OF ABSTRACT</b> SAR	